A latitudinal gradient of reference genomes

Ethan B. Linck[1]*, Carlos Daniel Cadena[2]

[1]*Department of Ecology, Montana State University*

[2]*Departamento de Ciencias Biológicas, Universidad de los Andes*

*Correspondence: ethan.linck@montana.edu

Abstract: Global inequality rooted in legacies of colonialism and uneven development can lead to systematic biases in scientific knowledge. In ecology and evolutionary biology, findings, funding and research effort are disproportionately concentrated at high latitudes while biological diversity is concentrated at low latitudes. This discrepancy may have a particular influence in fields like phylogeography, molecular ecology and conservation genetics, where the rise of genomics has increased the cost and technical expertise required to apply state-of-the-art methods. Here we ask whether a fundamental biogeographic pattern—the latitudinal gradient of species richness in tetrapods—is reflected in available reference genomes, an important data resource for various applications of molecular tools for biodiversity research and conservation. We also ask whether sequencing approaches differ between Global South and Global North, reviewing the last five years of conservation genetics research in four leading journals. We find that extant reference genomes are scarce relative to species richness at low latitudes, and that reduced-representation and whole-genome sequencing are disproportionately applied to taxa in the Global North. We conclude with recommendations to close this gap and improve international collaborations in biodiversity genomics.

Keywords: biodiversity genomics, conservation genetics, tetrapods, macroecology

Introduction: Scientific knowledge reflects both reality and the economic and social conditions that shape the practice of research (Hull 1990; Logino 2002; Stephan 2012). The coincidence of the Scientific Revolution and imperial expansion in Western Europe in the 16th through 18th centuries led to the accumulation and concentration of global capital in institutions of higher learning in the United Kingdom, France, Spain, the Netherlands, and Germany, funding foundational investigations in the natural sciences. Industrialization and the rise of capitalism in the 19th century stoked the boiler of the research endeavor, permitting the continued expansion of universities and birthing a learned aristocratic class with disposable time and income to commit to natural history (Opitz 2004; Opitz 2006). This legacy persists to this day: Research expenditures concentrated in a handful of majority Anglophone countries (May 1997), which in turn produce a disproportionate amount of scholarship (May 1997; King 2004) and often drive citation networks (Pasterkamp et al. 2007; Meneghini et al. 2008).

Ecology and evolutionary biology (EEB) faces a unique challenge in light of global scientific inequality, as while funding and research effort are disproportionately concentrated at high latitudes (Melles et al. 2019), biological diversity is concentrated at low latitudes (Willig et al. 2003). The consequences of this discrepancy are many, including systematic biases in research effort (Titley et al. 2017) and taxonomy (Freeman & Pennell 2021), major gaps in our understanding of natural history and species distributions (Collen et al. 2008; Feely & Sillman 2011), and a blinkered view of patterns of diversity and diversification across the tree of life (Reddy 2014; Cornwell et al. 2019). Increased research attention on tropical ecosystems has therefore been a stated priority for decades (Collen et al. 2008; Nori et al. 2020), but despite some promising strides towards international collaborations (Perez et al. 2018), partners in the Global North frequently set research agendas (Bradeley 2008; Asase et al. 2022), leading to

47 accusations of parachute science and persistent gaps in data quantity, quality, and type (Stocks et
48 al. 2008; Soares et al. 2023).

49 EEB disciplines where data collection and analysis are equipment- and cost-intensive are
50 especially likely to show increased discrepancies between the Global North and Global South. In
51 the 1960s and 70s, assays of molecular genetic markers in nonmodel organisms laid the
52 groundwork for phylogenetic systematics to shift away from a traditional focus on morphology,
53 increasing statistical power and requiring increasingly expensive tools and specialized skills
54 (Hillis et al. 1996). Concurrently, the rise of phylogeography brought an evolutionary
55 perspective to population biology, altering views on the fundamental units of management and
56 conservation (Avise 2000). In the late 2000s and early 2010s, reduced representation high-
57 throughput sequencing using library preparation methods such as RADseq and target-capture
58 began to proliferate as an alternative to traditional Sanger sequencing (Eklbom et al. 2011;
59 Lemmon & Lemmon 2013), boasting the advantages of rapidly generating much larger datasets
60 and requiring little to no prior information about a focal taxon's genome.

61 A third revolution is currently underway, as low-coverage whole genome resequencing
62 (hereafter WGS) begins to supplant reduced representation methods in some fields and taxa
63 (Ellegren et al. 2014; Toews et al. 2016). Unlike RADseq, target capture, and related approaches,
64 WGS typically requires a high-coverage reference genome with which to align samples. Because
65 sequencing an individual at sufficient depth remains costly—and because assembling such large
66 data is computationally intensive—reference genomes remain out of reach for many labs and, by
67 extension, species.

68 The problem may be particularly acute in conservation biology where decades of concern
69 about the so-called "conservation genetics gap" have highlighted a persistent disconnect between

70 the importance managers place on genetic information and its actual use in decision making
71 (Taylor et al. 2017). Hypotheses to explain the conservation genetics gap include skepticism
72 about its importance, the specialized knowledge required for analysis and interpretation, and cost
73 (Hoban et al. 2013). At low-resource institutions in the tropics a broader shift towards WGS
74 approaches may further discourage the development of the field in the very regions where the
75 greatest number of critically endangered species are found (Vamosi & Vamosi 2008; Bertola et
76 al. 2024), especially if technological advances become publication requirements at high impact
77 and broadly read journals.

78 Here we ask whether a fundamental biogeographic pattern—the latitudinal gradient of
79 species richness in tetrapods—is reflected in available reference genomes, an important data
80 resource for various applications of molecular tools for biodiversity research and conservation.
81 We hypothesized that inequities in economic development and access to scientific resources
82 would lead the number of species with assembled genomes to be greatest in the temperate zone,
83 not the tropics. We also ask whether sequencing approaches differ between the Global South and
84 Global North, reviewing the last five years of conservation genetics research in four leading
85 journals. We conclude by discussing strategies to improve international collaborations in
86 biodiversity genomics and boost the representation of species from the Global South in sequence
87 databases.

88

89

90 **Methods:**

91 *Reference genomes, species richness, and latitude.* We used the NCBI Datasets
92 command-line tools v.16.19.0 to download taxonomy metadata for the subset of species with an

assembled reference genome in the following taxa: birds (Class: Aves), mammals (Class: Mammalia), squamates (Order: Squamata), amphibians (Class: Amphibia), turtles (Order: Testudines), crocodilians (Order: Crocodilia) and tuataras (Order: Rhynchocephalia). We selected these groups—together comprising extant tetrapods—to provide a snapshot of animal diversity in relatively well-studied clades with different ecologies and evolutionary histories, while restricting the total dataset to a computationally manageable size. From this initial list we retained species with an exact match to GBIF's Backbone Taxonomy using rgbif v.3.8.0, and downloaded all observations of each backed by georeferenced voucher specimens in natural history museum collections (NHCs), excluding those without coordinates and those flagged for geospatial issues. We repeated this process for all species in each higher-level taxon represented in our list of reference genomes (i.e., downloaded metadata for all georeferenced tetrapod specimens on GBIF).

Filtering these aggregated datasets to contain only species with 10 or more records, we generated convex hull polygons for each as an approximation of their geographic distribution. Overlaying these on a shapefile of Earth's landmasses from rnaturalearth v.1.0.1, we calculated species richness as the number of overlapping convex hulls in 2-degree x 2-degree grid cells, standardizing this value by subtracting mean global species richness and dividing by its standard deviation. We subtracted the number of species with reference genomes from total species richness to determine the regions with the largest representation gap in genomic resources, again standardizing the difference. To evaluate the significance and slope of a correlation between species richness and the absolute value of latitude, we performed simple linear regressions in R v.4.4.0, analyzing species with reference genomes and our full dataset separately.

Literature review. To evaluate how the geography of authorship might impact sequencing strategy of studies in conservation biology, we performed a restricted Web of Science literature search on 29 June 2024 for conservation genetics papers published in the last five years in the journals *Conservation Genetics*, *Molecular Ecology*, *Journal of Heredity*, and *Conservation Biology*. We used the queries 'SO="Conservation Genetics"' and 'SO=("Molecular Ecology" OR "Journal of Heredity" OR "Conservation Biology") AND (TS="Conservation Genet*" OR KP="Conservation Genet*" OR TI="Conservation Genet*"'), excluding reviews, genome announcements, meta-analyses, preprints, and studies that were purely simulations.

We then manually reviewed each study, first assigning the home institution of the first and last author to the Global North or Global South using the 2024 UN Trade and Development Classifications. We categorized its sequencing approach as reduced representation, WGS, Sanger sequencing, microsatellites, or other, and described its overall focus using tiered categories based on discussion in Bertola et al. 2024. These tiers were: 1) Taxonomy / systematics, identification, or sexing; 2) Phylogeography / population genetic structure, estimating genetic diversity, and inferring demographic history; and 3) Detecting outlier loci, quantifying runs of homozygosity, and evaluating adaptive potential. When studies employed more than one sequencing approach or addressed goals belonging to multiple tiers, we assigned them to a single category on the basis of their most data-intensive method or question.

Results:

Sampled Gradients of Species Richness. Our list of tetrapod reference genomes from NCBI included 1159 bird species, 795 mammal species, 123 amphibian species, 39 turtle

species, 6 crocodile species, and the tuatara (*Sphenodon punctuatus*). This total represented 6.8% of the 30,832 tetrapod species with a georeferenced preserved specimen record on GBIF. Species with a reference genome were associated with 3,048,136 specimens, or 40.7% of the 7,478,867 tetrapod specimens with available metadata. Following filtering out species with 10 or fewer observations, we retained 1,859 species with a reference genome, or 8.6% of the 21,583 tetrapod species meeting the same criteria.

Minimum tetrapod species richness in both datasets was 1. Maximum species richness calculated from the reference genome dataset was 705, occurring in a grid cell centered on south Florida, USA (**Figure 1C**). Maximum species richness in the full dataset was 2698, occurring in a grid cell centered on the western Amazon basin and east slope of the Andes in Ecuador near the Peruvian border (**Figure 1C**); this was also where the greatest gap between sequenced species and total species richness was observed (**Figure 1A**).

Species richness was negatively related to the absolute value latitude in both regressions, albeit with a much steeper slope when data from all tetrapod species were included (reference genomes only, $\beta = -4.892$, adjusted $R^2=0.6439$; $p<0.001$; full data, $\beta = -18.65$, adjusted $R^2=0.7856$ $p<0.001$) (**Figure 1B**). Because data visualization indicated there might be a distinct breakpoint in the relationship at mid latitudes—potentially reflecting a transition from the influence of sampling effort effects to true biogeographic signal—we fit an additional piecewise linear regression model with the R package segmented v.2.1-0. This model identified a breakpoint at 39.819, fitting a segment with a slope of $\beta = -0.312$ before it, and a segment with a slope of $\beta = -7.3634$ after it ($p=0.0408$; adjusted $R^2=0.7184$).

160 *Literature Review.* After excluding papers that did not meet our stated criteria, we
161 reviewed 394 empirical conservation genetics articles published between 1 January 2020 and 29
162 June 2024. This list included 248 papers from the journal *Conservation Genetics,* 11 papers from
163 *Journal of Heredity*, and 35 articles from *Molecular Ecology;* no appropriate studies from
164 *Conservation Biology* were identified. Of these, 62 included a first or senior author from the
165 Global South, while 342 included a first or senior author from the Global North. Ninety-eight
166 included sampling from a focal taxon or focal taxa in the Global South, with 277 sampling a
167 taxon or taxa from the Global North. Microsatellites were the most used sequencing strategy
168 among Global South authors and in studies of Global South taxa, while reduced representation
169 genome sequencing approaches were most common in the Global North. For authors and taxa in
170 both the Global South and the Global North, Tier 2 studies (phylogeography / population genetic
171 structure, estimating genetic diversity, or inferring demographic history) were most common.
172 Further details are provided in **Table 1** and **Table 2**.
173
174 **Discussion:** Extant reference genomes fail to reflect the overwhelming concentration of
175 tetrapod species richness in the tropics and are strongly biased towards species at mid-latitudes in
176 the Northern Hemisphere (**Figure 1**). This pattern is almost certainly a result of global
177 inequalities in economic development and its resulting effects on research productivity (May
178 1997; King 2004). Its consequences will likely include increasing an already profound
179 methodological gap in sequencing approaches between molecular ecologists in the Global North
180 and the Global South (**Table 1**).
181 Our analysis contrasts patterns inferred from both a traditional source of biodiversity
182 data—vouchered specimens in NHCs—and a contemporary genomic resource archive, the NIH

183 NCBI. Our description of the latitudinal gradient of total tetrapod species richness is broadly
184 similar to other recent macroecological studies of the phenomenon (Roll et al. 2017; Quintero et
185 al. 2023), with global hotspots concentrated in northwest South America and Central Africa
186 (**Figure 1C**) and an approximately monotonic decline in latitudinal species richness maxima
187 from the equator to the poles. In contrast, the latitudinal gradient of richness of tetrapod species
188 with reference genomes is flattened, showing only moderate declines at high latitudes and a mid-
189 latitude peak in species richness in the Northern Hemisphere (**Figure 1B**).

190 This difference is especially notable because we made no effort to correct for disparities
191 in historical specimen collection across latitude, with the consequence that our 'true' species
192 richness gradient significantly underestimates biodiversity in the tropics. Across longitude, our
193 analysis appears to underestimate diversity in East Asia, Indonesia, and Oceania (Quintero et al.
194 2023), likely due to both the coarse grain of our study and the region's greater distance from the
195 large NHCs in Europe and North America that are the backbone of curated GBIF data.
196 Regardless, species with publicly available reference genomes as of July 2024 are more
197 reflective of socioeconomic conditions than biogeographic reality.

198 If both natural history collections and contemporary bioinformatics resources reflect
199 historical inequalities in development and scientific capacity, why do data from the former better
200 approximate the latitudinal gradient of species richness? Part of the answer lies in their different
201 goals: while the mission of many NHCs is to explicitly catalog and archive regional or global
202 biodiversity, the NCBI Genome Browser is often used as a repository for open data publication
203 requirements and is less often an end unto itself. Another part of the answer lies in shifting global
204 politics: as the birth and golden age of NHCs coincided with the heyday of Western colonialism

205 (De Vos 2007; Quintero Toro 2012), access to tropical habitat by collectors was less restricted by
206 concerns of either sovereignty or Indigenous land tenure.

207 We are not naive enough to believe that scientific colonialism is no longer a problem
208 (Asase et al. 2022; Soares et al. 2023). However, we suggest that a combination of evolving
209 norms and persistent obstacles of cost have led many relatively well-resourced scientists in the
210 Global North to prioritize generating reference genomes for local taxa. In spite of rapid declines
211 in the per base pair cost of whole genome sequencing (Lou et al. 2021), high-coverage
212 sequencing remains a significant expense: while averages are hard to come by in this
213 increasingly privatized sector, one of us (E.L.) recently paid ~$13,000 2024 USD for long-read
214 and HiC sequencing of a North American passerine bird. Even in North America, this figure
215 likely pushes small, single PI labs to prioritize investing in generating resources for species at the
216 center of their research program, or otherwise likely to provide long-term utility—which are
217 often those in their own backyards.

218 In the spirit of the traditional mission of NHCs, the past decade has seen several
219 interrelated, international initiatives to increase taxonomic diversity in high-quality reference
220 genomes (e.g. O'Brien et al. 2014; Koepfli et al. 2015; Cheng et al. 2018; Rhie et al. 2021).
221 These collaborations have had a profound impact on biodiversity genomics, helping to close
222 what was surely an even larger gap in empirical patterns of species richness and species richness
223 represented by NCBI. In some cases, they have provided researchers with early access to draft
224 assemblies of nonmodel organisms (e.g., Linck et al. 2020) or with support and resources to
225 produce new assemblies (e.g., Cadena et al. 2024), a scenario that suggests the resource gap may
226 be slightly less dire in practice than reported here. Yet in a world of limited time, finite
227 resources, and incompletely described biodiversity, sequencing at scale is not immune to its own

228 biases. For example, a recent publication introducing an attempt to generate reference genomes
229 for all vertebrates (Rhie et al. 2021) included 127 authors affiliated with 102 institutions. Of
230 these, only 13 are in the Global South: 4 in China, 3 in Korea, 2 in Malaysia, 1 in Singapore, 1 in
231 Qatar, and 1 in Colombia. No authors or institutional affiliations from Africa, Indonesia, or
232 Oceania outside of Australia and New Zealand were included. Though understandable
233 considering the current distribution of scientists and resources—though not of species—the
234 imbalance seems likely to perpetuate representational biases in the near-term.

235 Current efforts by international collaborations like the Amphibian Genomics Consortium
236 to increase geographic representation and to offer support and opportunities to researchers from
237 developing countries and underrepresented groups (Kosch et al. 2024) are steps in the right
238 direction in the path to make the field of biodiversity genomics more equitable. In line with the
239 Convention on Biological Diversity's Nagoya Protocol (Secretariat of the Convention on
240 Biodiversity 2011), another critical dimension of the conversation about equitable generation of
241 encyclopedias of reference genomes is the need for researchers to build strong partnerships with
242 Indigenous peoples and other local communities, allowing them to participate in and benefit
243 from the different phases and products of sequencing projects (Ambler et al. 2020; Colella et al.
244 2023; Mc Cartney et al. 2023).

245 If tropical species are underrepresented on NCBI, we would expect that they are only
246 rarely studied using whole-genome resequencing (and other sequencing strategies dependent on
247 a reference genome). Our review of conservation genetics papers published in *Molecular*
248 *Ecology, Journal of Heredity, Conservation Genetics,* and *Conservation Biology* over the last
249 five years suggests this is indeed the case. While the Global South / Global North binary and
250 measures of human development more generally are only imperfectly correlated with latitude,

251 their association nonetheless indicates that WGS is only rarely applied in conservation genetics
252 studies in the tropics (n=6), and almost never by a leading researcher with a primary affiliation to
253 a research institution in the region (n=1) (**Table 1**).

254 Similarly, for Global South scientists and for Global South focal taxa, microsatellites
255 remain the most common molecular approach; for scientists and focal taxa in the Global North,
256 reduced representation approaches dominate. We believe this is reflective of the expense and
257 limited availability of high-throughput sequencing in the Global South, regardless of whether
258 reads are assembled *de novo* or aligned to a reference. Lastly, we point out that across all
259 categories, the vast majority of scientists (87%) and species (77%) in our sample originate in the
260 Global North. Though our choice of North American or European-based, English language
261 journals precludes generalization, it nevertheless seems safe to interpret this as indicating the
262 field of conservation genetics—let alone conservation genomics—is in its infancy in the tropics.
263 Setting appropriate goals, targets, and indicators to effectively conserve and monitor global
264 genetic diversity will require this situation to be remedied (Hoban et al. 2021).

265 We highlight discrepancies between available reference genomes and global
266 biogeographic patterns to encourage increased, equitable collaboration between scientists in the
267 Global North and Global South. In light of this, we make three simple recommendations (see
268 also Bertola et al. 2023). First, we encourage scientists from resource-rich institutions to consider
269 allocating effort and funds towards generating reference genomes that serve the needs of
270 managers and researchers in the Global South. Second, we support the continued development of
271 multinational sequencing projects, but ask funders and senior personnel to increasingly consider
272 prioritization, inclusion, and capacity building in areas of the world with rich biodiversity and
273 limited resources to study it using genomic tools. Third, we ask journals to consider issues of

access and cost in editorial guidelines and decisions: While high-throughput sequencing is increasingly expected by editorial boards and reviewers at high-impact journals, it is not essential to address a variety of research questions (Bertola et al. 2023) and remains out of reach for most scientists residing where most of the world's species occur. If ecology, evolution, and conservation aim to accurately catalog and effectively protect life on Earth, remedying inequalities in genomic resources should be a major priority.

Data Accessibility: DOIs for GBIF downloads, processed datasets, and a digital notebook containing code to perform these analyses and generate Figure 1 are available at https://github.com/elinck/lat_grad_genome and from Data Dryad (pending).

Acknowledgments: We thank Marty Kardos for the invitation to participate in this special issue.

References:

Ambler, J., Dearden, P. K., Wilcox, P., Hudson, M., & Tiffin, N. (2021). Including digital sequence data in the Nagoya Protocol can promote data sharing. Trends in biotechnology, 39(2), 116-125.

Asase, A., Mzumara-Gawa, T. I., Owino, J. O., Peterson, A. T., & Saupe, E. (2022). Replacing "parachute science" with "global science" in ecology and conservation biology. Conservation Science and Practice, 4(5), e517.

Avise, J. C. (2000). Phylogeography: the history and formation of species. Harvard university press.

Bertola, L. D., Brüniche-Olsen, A., Kershaw, F., Russo, I. R. M., MacDonald, A. J., Sunnucks, P., ... & Segelbacher, G. (2024). A pragmatic approach for integrating molecular tools into biodiversity conservation. Conservation Science and Practice, 6(1), e13053.

Bradley, M. (2008). On the agenda: North–South research partnerships and agenda-setting processes. Development in Practice, 18(6), 673-685.

Cadena, C. D., Pabón, L., DoNascimiento, C., Abueg, L., Tilley, T., O-Toole, B., ... & Torres, M. (2024). A reference genome for the Andean cavefish Trichomycterus rosablanca (Siluriformes, Trichomycteridae): Building genomic resources to study evolution in cave environments. Journal of Heredity, 115(3), 311-316.

Colella, J. P., Silvestri, L., Súzan, G., Weksler, M., Cook, J. A., & Lessa, E. P. (2023). Engaging with the Nagoya Protocol on Access and Benefit-Sharing: recommendations for noncommercial biodiversity researchers. Journal of Mammalogy, 104(3), 430-443.

Collen, B., Ram, M., Zamin, T., & McRae, L. (2008). The tropical biodiversity data gap: addressing disparity in global monitoring. Tropical Conservation Science, 1(2), 75-88.

Cornwell, W. K., Pearse, W. D., Dalrymple, R. L., & Zanne, A. E. (2019). What we (don't) know about global plant diversity. Ecography, 42(11), 1819-1831.

Cheng, S., Melkonian, M., Smith, S. A., Brockington, S., Archibald, J. M., Delaux, P. M., ... & Wong, G. K. S. (2018). 10KP: A phylodiverse genome sequencing plan. Gigascience, 7(3), giy013.

De Vos, P. S. (2007). Natural history and the pursuit of empire in eighteenth-century Spain. Eighteenth-Century Studies, 40(2), 209-239.

Ekblom, R., & Galindo, J. (2011). Applications of next generation sequencing in molecular ecology of non-model organisms. Heredity, 107(1), 1-15.

Ellegren, H. (2014). Genome sequencing and population genomics in non-model organisms. Trends in Ecology & Evolution, 29(1), 51-63.

Feeley, K. J., & Silman, M. R. (2011). The data void in modeling current and future distributions of tropical species. Global Change Biology, 17(1), 626-630.

Freeman, B. G., & Pennell, M. W. (2021). The latitudinal taxonomy gradient. Trends in Ecology & Evolution, 36(9), 778-786.

Hillis, D. M., Moritz, C., & Mable, B. K. (1996). Molecular systematics (Vol. 23). Sinauer.

Hoban, S. M., Hauffe, H. C., Pérez-Espona, S., Arntzen, J. W., Bertorelle, G., Bryja, J., ... & Bruford, M. W. (2013). Bringing genetic diversity to the forefront of conservation policy and management. Conservation Genetics Resources, 5, 593-598.

Hoban, S., Bruford, M. W., Funk, W. C., Galbusera, P., Griffith, M. P., Grueber, C. E., ... & Vernesi, C. (2021). Global commitments to conserving and monitoring genetic diversity are now necessary and feasible. Bioscience, 71(9), 964-976.

Hull, D. L. (1990). Science as a process: an evolutionary account of the social and conceptual development of science. University of Chicago Press.

King, D. A. (2004). The scientific impact of nations. Nature, 430(6997), 311-316.

Koepfli, K. P., Paten, B., Genome 10K Community of Scientists, & O'Brien, S. J. (2015). The Genome 10K Project: a way forward. Annu. Rev. Anim. Biosci., 3(1), 57-111.

Kosch, T. A., Torres-Sanchez, M., Liedtke, H. C., Summers, K., Yun, M. H., Crawford, A. J., ... & Amphibian Genomics Consortium (AGC). (2024). The Amphibian Genomics Consortium: advancing genomic and genetic resources for amphibian research and conservation. bioRxiv, 2024-06.

Lemmon, E. M., & Lemmon, A. R. (2013). High-throughput genomic data in systematics and phylogenetics. Annual Review of Ecology, Evolution, and Systematics, 44, 99-121.

Linck, E., Freeman, B. G., & Dumbacher, J. P. (2020). Speciation and gene flow across an elevational gradient in New Guinea kingfishers. Journal of Evolutionary Biology, 33(11), 1643-1652.

Logino, H. (2019). The Social Dimensions of Scientific Knowledge. In E. N. Zalta (Ed.), The Stanford Encylopedia of Philosophy (Summer 2019 Edition). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/sum2019/entries/scientific-knowledge-social/

Lou, R. N., Jacobs, A., Wilder, A. P., & Therkildsen, N. O. (2021). A beginner's guide to low-coverage whole genome sequencing for population genomics. Molecular Ecology, 30(23), 5966-5993.

May, R. M. (1997). The scientific wealth of nations. Science, 275(5301), 793-796.

Mc Cartney, A. M., Head, M. A., Tsosie, K. S., Sterner, B., Glass, J. R., Paez, S., ... & Hudson, M. (2023). Indigenous peoples and local communities as partners in the sequencing of global eukaryotic biodiversity. npj Biodiversity, 2(1), 8.

Melles, S. J., Scarpone, C., Julien, A., Robertson, J., Levieva, J. B., Carrier, C., ... & Morales, K. (2019). Diversity of practitioners publishing in five leading international journals of applied ecology and conservation biology, 1987–2015 relative to global biodiversity hotspots. Ecoscience, 26(4), 323-340.

Meneghini, R., Packer, A. L., & Nassi-Calo, L. (2008). Articles by Latin American authors in prestigious journals have fewer citations. PloS ONE, 3(11), e3804.

Nori, J., Loyola, R., & Villalobos, F. (2020). Priority areas for conservation of and research focused on terrestrial vertebrates. Conservation Biology, 34(5), 1281-1291.

OBrien, S. J., Haussler, D., & Ryder, O. (2014). The birds of Genome10K. GigaScience, 3(1), 2047-217X.

Opitz, D. L. (2004). Aristocrats and professionals: Country-house science in late-Victorian Britain. University of Minnesota.

Opitz, D. L. (2006). " This House is a Temple of Research": Country-House Centres for Late-Victorian Science.

Pasterkamp, G., Rotmans, J., de Kleijn, D., & Borst, C. (2007). Citation frequency: A biased measure of research impact significantly influenced by the geographical origin of research articles. Scientometrics, 70(1), 153-165.

Perez, T. M., & Hogan, J. A. (2018). The changing nature of collaboration in tropical ecology and conservation. Biotropica, 50(4), 563-567.

Quintero, I., Landis, M. J., Jetz, W., & Morlon, H. (2023). The build-up of the present-day tropical diversity of tetrapods. Proceedings of the National Academy of Sciences, 120(20), e2220672120.

Quintero Toro, C. (2012). Birds of empire, birds of nation: A history of science, economy, and conservation in United States-Colombia relations. Ediciones Uniandes-Universidad de los Andes.

Reddy, S. (2014). What’s missing from avian global diversification analyses?. Molecular Phylogenetics and Evolution, 77, 159-165.

Rhie, A., McCarthy, S. A., Fedrigo, O., Damas, J., Formenti, G., Koren, S., ... & Jarvis, E. D. (2021). Towards complete and error-free genome assemblies of all vertebrate species.

Nature, 592(7856), 737-746.

Roll, U., Feldman, A., Novosolov, M., Allison, A., Bauer, A. M., Bernard, R., ... & Meiri, S. (2017). The global distribution of tetrapods reveals a need for targeted reptile conservation. Nature Ecology & Evolution, 1(11), 1677-1682.

Secretariat of the Convention on Biodiversity (2011). Nagoya Protocol on Access to Genetic Resources and the Fair and Equitable Sharing of Benefits Arising from their Utilization to the Convention on Biological Diversity : text and annex. Montreal, Canada, Secretariat of the Convention on Biodiversity, 15pp. DOI: http://dx.doi.org/10.25607/OBP-789

Seth, S. (2009). Putting knowledge in its place: science, colonialism, and the postcolonial. Postcolonial studies, 12(4), 373-388.

Soares, L., Cockle, K. L., Ruelas Inzunza, E., Ibarra, J. T., Miño, C. I., Zuluaga, S., ... & Martins, P. V. R. (2023). Neotropical ornithology: Reckoning with historical assumptions, removing systemic barriers, and reimagining the future. Ornithological Applications, 125(1), duac046.

Stocks, G., Seales, L., Paniagua, F., Maehr, E., & Bruna, E. M. (2008). The geographical and institutional distribution of ecological research in the tropics. Biotropica, 40(4), 397-404.

Stephan, P. (2012). How economics shapes science. Harvard University Press.

Taylor, H. R., Dussex, N., & van Heezik, Y. (2017). Bridging the conservation genetics gap by identifying barriers to implementation for conservation practitioners. Global Ecology and Conservation, 10, 231-242.

Titley, M. A., Snaddon, J. L., & Turner, E. C. (2017). Scientific research on animal biodiversity is systematically biased towards vertebrates and temperate regions. PloS one, 12(12), e0189577.

Toews, D. P., Campagna, L., Taylor, S. A., Balakrishnan, C. N., Baldassarre, D. T., Deane-Coe, P. E., ... & Winger, B. M. (2016). Genomic approaches to understanding population divergence and speciation in birds. The Auk: Ornithological Advances, 133(1), 13-30.

Willig, M. R., Kaufman, D. M., & Stevens, R. D. (2003). Latitudinal gradients of biodiversity: pattern, process, scale, and synthesis. Annual review of ecology, evolution, and systematics, 34(1), 273-309.

Vamosi, J. C., & Vamosi, S. M. (2008). Extinction risk escalates in the tropics. PLoS One, 3(12), e3886.

Figures & Tables:

Table 1. Summary of regional authorship affiliation, sequencing strategy, and sampled focal species range for empirical conservation genetics papers from 2019-2024 in four leading journals. Integers indicate the total number of studies in each category, while numbers in parentheses refer to its proportion out of all reviewed articles (n=394). Papers were assigned to a sequencing strategy based on the most data-intensive approach they employed (i.e., a study applying both Sanger sequencing and microsatellites would be assigned to the 'Microsatellites' category.)

Sequencing Strategy	**Global South Author**	**Global North Author**	**Global South Taxon**	**Global North Taxon**
Sanger	13 (0.0329)	36 (0.0913)	21 (0.0532)	21 (0.0532)
Microsatellites	28 (0.0710)	123 (0.3121)	40 (0.1015)	103 (0.2614)
Reduced Representation	15 (0.0381)	146 (0.3706)	27 (0.0685)	127 (0.3223)
WGS	2 (0.0050)	27 (0.0685)	6 (0.0152)	18 (0.0456)
Other	4 (0.0101)	10 (0.0254)	4 (0.0101)	8 (0.0203)

449 **Table 2.** Summary of regional authorship affiliation, study goals, and sampled focal
450 species range among reviewed papers. Study goals refer to broad tiers of research questions with
451 increasing data requirements. Papers were assigned to each on the basis of the most data-
452 intensive analysis they employed (i.e., a paper inferring population genetic structure and
453 identifying loci under selection would be assigned to the tier 3).
454

Study Goals	**Global South Author**	**Global North Author**	**Global South Taxon**	**Global North Taxon**
1. Taxonomy / systematics, identification, or sexing	6 (0.0152)	25 (0.0635)	11 (0.0279)	17 (0.04315)
2. Phylogeography / population genetic structure, estimating genetic diversity, and inferring demographic history	61 (0.1548)	285 (0.7234)	102 (0.2588)	248 (0.6294)
3. Detecting outlier loci, quantifying runs of homozygosity, and evaluating adaptive potential	1 (0.0025)	35 (0.0888)	4 (0.0102)	31 (0.0787)

455
456

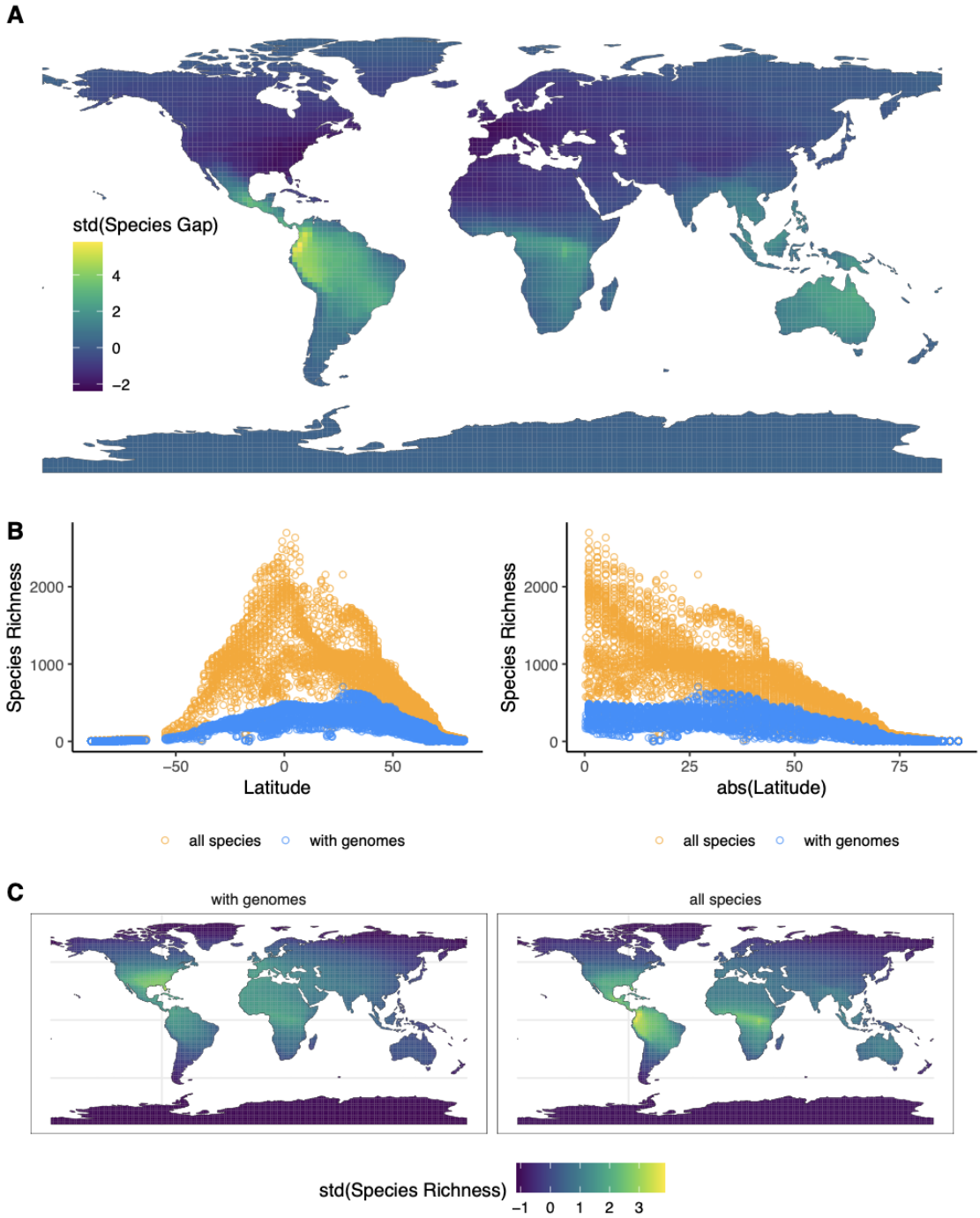

Figure 1. A) Tropical species are underrepresented among available reference genomes. Colors reflect the standardized difference between total species richness and the number of species with an assembled reference genome on the NCBI genome browser. B) Richness among species with reference genomes does not reflect global patterns of biodiversity. Blue circles represent total local species richness in 2-degree by 2-degree grid cells, while gold circles represent richness of species with assembled reference genomes. C) Global patterns of species richness calculated from species with reference genomes and all species with NHC specimen records on GBIF, respectively.